

OPTIMIZACIÓN DE MALLA VARIABLE APLICADA AL APRENDIZAJE DE CLASIFICADORES BAYESIANOS

Byron Oviedo Bayas^{1,*}, Erika Zamora Cevallos², Amilkar Puris Cáceres³, Cristian Zambrano-Vega⁴, Jorge Gómez Gómez⁵

^{1,2,3,4} Quevedo State Technical University, UTEQ, Quevedo-Los Ríos-Ecuador, 5 Cordoba University, Montería-Colombia

*Corresponding Author: boviedo@uteq.edu.ec

Resumen: En este artículo se define una forma de aprendizaje estructural para clasificadores bayesianos utilizando la Meta heurística Optimización basada en Mallas Variables, VMO. El método se basa en encontrar la topología de redes bayesianas arbitrarias que mejor clasifique los datos. En el proceso se utiliza una técnica envolvente para la tarea de aprendizaje supervisado. Este problema de optimización es complejo, ya que el espacio crece de manera exponencial en dependencia del número de variables. La propuesta fue probada en un escenario educativo y comparada con otros clasificadores bayesianos utilizando software de acceso libre como Elvira y Weca.

Palabras Clave: Redes bayesianas, entrenamiento estructural, Meta heurística, malla variable, optimización.

OPTIMIZATION OF VARIABLE MESH APPLIED TO THE LEARNING OF BAYESIAN CLASSIFIERS

Abstract: : This paper defines a form of structural learning for Bayesian classifiers using the heuristic Meta Optimization based on Variable Meshes, VMO. The method is based on finding the topology of arbitrary Bayesian networks that best classify the data. In the process, a wraparound technique is used for the supervised learning task. This optimization problem is complex, since space grows exponentially depending on the number of variables. The proposal was tested in an educational setting and compared with other Bayesian classifiers using free access software such as Elvira and Weca.

Key words: Bayesian networks, structural training, heuristic goal, variable mesh, optimization.

I INTRODUCCIÓN

El aprendizaje estructural de clasificadores Bayesianos es una tarea compleja que amerita mucho costo computacional [1]. En la actualidad este proceso se puede realizar a partir de dos enfoques fundamentales:

- Métodos de filtrado.- Estos métodos se basan en el uso de medidas de asociación o métricas que se calculan en los datos. El clasificador se construye en base a estas medidas, por ejemplo, optimizando una métrica como se hace en la construcción del Clasificador Bayesiano Simple Aumentado con un Árbol TAN [2]. Hay dos enfoques principales dentro del aprendizaje de clasificadores bayesianos a partir de un conjunto de observaciones: pruebas de independencia [3] y métodos de score + búsqueda [4]. La primera, está basada en hacer test estadísticos relacionados con los datos que están representados en el grafo. En el segundo existe una métrica que permite medir la idoneidad de un grafo dado un conjunto de datos propuestos.

- Métodos envolventes.- El mejor clasificador se calcula optimizando su comportamiento en los datos de entrenamiento usando un método de validación cruzada [5] por ejemplo usando la validación basada en 10 tandas.

En nuestro caso vamos a usar un método envolvente para el aprendizaje de los clasificadores, donde se tiene como problema el encontrar la topología que mejor clasifique los datos. De hecho el problema de aprender una red bayesiana que optimice una métrica como Bayesian Dirichlet likelihood-equivalence uniform joint distribution BDeu es un problema de tiempo polinomial no determinista NP-difícil [1]. Por ese motivo, es conveniente usar métodos de optimización combinatoria que permitan obtener buenas soluciones en tiempos razonables.

Entre los métodos de optimización más utilizados para el aprendizaje se encuentran las meta heurísticas poblacionales, definidos como algoritmos iterativos de búsqueda en el espacio de soluciones S que emplean un conjunto de soluciones (población) en cada iteración del algoritmo para optimizar una función objetivo [6].

En los últimos años ha habido un crecimiento espectacular en el desarrollo de las meta heurísticas poblacionales. Un ejemplo de esto es la meta heurística Optimización basada en Mallas Variables (VMO). Dicho método utiliza una malla de punto en el espacio n dimensional que se expande y se contrae para recorrer el espacio de búsqueda [7]. Esta meta heurística ha sido presentada tanto en problemas discretos como continuos [8], [9] entre otros.

La propuesta de aprendizaje estructural fue incor-

porada en el software Elvira para realizar las pruebas correspondientes.

II DESARROLLO

El estudio empieza analizando brevemente los problemas de optimización y se determina la importancia de considerar a los operadores que permitirán mover a la población a las zonas favorables del espacio de solución. Luego se describe de manera general la meta heurística optimización de malla variable, explicando cómo genera la malla inicial, los nodos en dirección a los extremos locales y al extremo global, así como la generación de los nodos a partir de las fronteras de la malla. Luego se analizan características del método y sus operaciones. Finalmente se aplica el método para estimar clasificadores bayesianos y se realizan experimentos comparando con otros clasificadores. Vamos a usar un método de envolvente y la función que se optimiza es el comportamiento en el conjunto de entrenamiento mediante validación cruzada. Algunos algoritmos de aprendizaje estructural de redes bayesianas con los que se van a comparar los resultados son ByNet, BayesChaid, BayesPSO [10].

DESCRIPCIÓN DE LA META HEURÍSTICA VMO

La optimización de malla variable (VMO), [7], es una meta heurística poblacional donde las soluciones son distribuidas como una malla en el espacio m -dimensional. La población está compuesta por p nodos o soluciones candidatas (n_1, n_2, \dots, n_p), codificadas como un vector de puntos flotantes de m componentes, $n_i = (v_{i1}, v_{i2}, \dots, v_{im})$.

Esta meta heurística aparece como un esfuerzo de utilizar de forma directa las soluciones más promisorias (extremos locales, extremo global y soluciones fronteras) para dirigir el proceso de búsqueda del modelo. Bajo este principio, la población puede converger de manera prematura a zonas ya identificadas como prometedoras y dejar de explorar otras zonas desconocidas. Para evitar esta situación desfavorable y propiciar la diversidad en la población, VMO implementa un mecanismo de limpieza adaptativa que filtra la población, manteniendo solamente los nodos más representativos de cada zona explorada. El componente adaptativo se adquiere a partir de la forma en que se calcula la distancia de separabilidad permitida en cada iteración del algoritmo [11].

De forma general el proceso de búsqueda desarrollado por VMO se realiza a partir de las siguientes ope-

raciones:

- Expansión: Proceso que se lleva a cabo en cada iteración para explorar las zonas representadas por las mejores soluciones obtenidas (extremos locales, el extremo global y las soluciones fronteras).

- Contracción: Mecanismo aplicado para reducir la población luego del proceso de expansión, manteniendo en la población los nodos más representativos.

Ambas operaciones son las encargadas de incorporar exploración y explotación del espacio de búsqueda.

APRENDIZAJE ESTRUCTURAL DE REDES BAYESIANAS

El aprendizaje estructural de una red bayesiana representa uno de los elementos fundamentales en el desempeño de un clasificador bayesiano. Este problema cae en la categoría NP-difícil [4], por lo que ha sido objeto de estudio de un sin número de investigaciones en el área.

El objetivo principal de esta fase es obtener una estructura de red bayesiana partiendo de bases de datos, es decir, se toma en consideración las relaciones de dependencia e independencia entre todas las variables. Las diferentes técnicas de aprendizaje estructural basan su procesamiento en el tipo de estructura o topología de la red (árboles, poli-árboles o redes múltiplemente conectadas).

Específicamente dentro de las redes múltiplemente conectadas aparecen dos enfoques diferentes de aprendizaje diferentes:

- Métodos que basan su funcionamiento en medidas de búsqueda y ajuste.

- Métodos que basan su funcionamiento en pruebas de independencia o también logran basarse en restricciones.

El método de búsqueda y ajuste propone el uso de una etapa de un proceso de búsqueda de estructuras y una fase que evalúa la calidad de la estructura obtenida permitiendo realizar ajustes en la topología de las redes [12], [13].

VMO APLICADO AL APRENDIZAJE ESTRUCTURAL DE REDES BAYESIANAS

En esta sección aplicaremos el algoritmo VMO al problema de aprendizaje estructural de un clasificador bayesiano. Para ello, se redefinen algunos operadores del algoritmo ajustándolos a las características del problema. Luego se presenta un estudio de los parámetros del modelo y por último se realiza un análisis compara-

tivo con otros clasificadores bayesianos del estado del arte, utilizando un conjunto de bases de conocimiento del repositorio de la UCI.

Modificaciones del algoritmo VMO

El primer paso en el proceso de adaptación del algoritmo VMO lo determina la representación de los nodos o soluciones. En este sentido tomaremos el nodo como un conjunto de arcos (X_i, X_j) , $i \neq j$ dirigidos, que representan las relaciones de dependencia entre las variables X_i y X_j de la red. La figura 1 presenta un ejemplo de un nodo para una red de 4 variables.

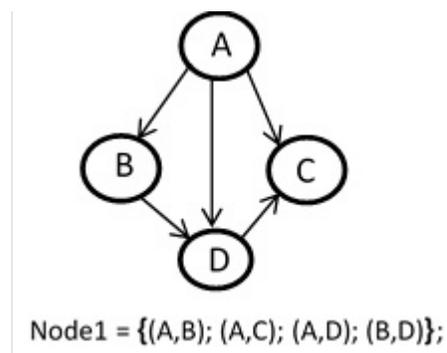


Figura 1. Representación de una Red Bayesiana como conjunto de arcos

Otro elemento que definiremos como se calcula la similitud entre dos redes. Nuestra propuesta radica en utilizar la operación de intersección entre dos conjuntos para obtener los arcos que comparten dos redes o nodos (\cap) . Luego la cardinalidad del conjunto solución, lo tomaremos como la similitud entre ambas redes (ecuación 1).

$$\text{Similitud}(BN_1, BN_2) = |BN_1 \cap BN_2| \quad (1)$$

Otro de los elementos que se definen en la propuesta es la forma de generar nuevas redes para el proceso de expansión de la malla. Es importante señalar que las redes que se generen no pueden tener ciclos. A continuación se definen las formas de generar los nuevos:

Generación de nodos al azar: este tipo de nodo se construye con arcos dirigidos generados al azar hasta que todas las variables estén conectadas.

Generación a partir de un nodo: esta generación se lleva a cabo en la función H del modelo VMO, para explorar las fronteras de la malla. Para ello, se selecciona aleatoriamente un arco del nodo y se procede a cambiar su orientación. El nuevo arco es incluido en la red sólo si no introduce ciclo.

Generación a partir de dos nodos: este caso se utiliza

$$BN^* = (BN_1 / BN_2) \cup (BN_2 / BN_1) \quad (2)$$

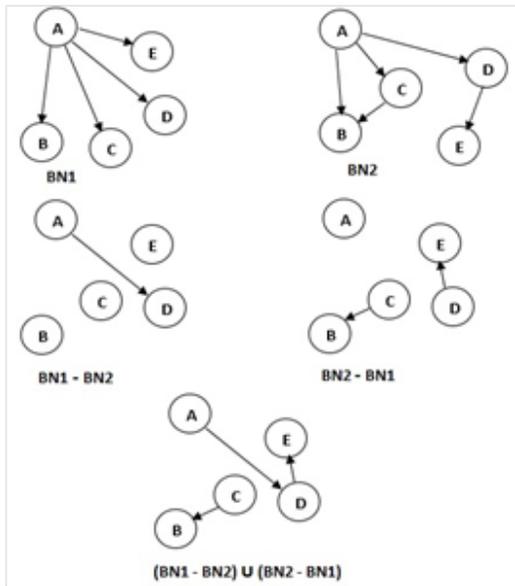


Figura 2. Operación unión de las diferencias entre BN1 y BN2

En el proceso de creación de las nuevas redes, se puede dar el caso de que alguna nueva red contenga ciclos, en este caso se aplica un algoritmo de eliminación de ciclos, que selecciona un arco de forma aleatoria entre, el conjunto de arcos que forman el ciclo, y se invierte el sentido del arco. Este proceso se repite hasta que se eliminen todos los ciclos de la red.

Cota de distancia o valor de separabilidad entre soluciones: es otro elemento que se redefinió en la propuesta (ver ecuación 3) donde A representa el número de variables o características del problema, y al igual que la propuesta original, c y C representan las evaluaciones de la función objetivo (actual y total respectivamente).

$$\xi = \begin{cases} 0,8 * A, & c < C / 4 \\ 0,7 * A, & c < C / 2 \\ 0,6 * A, & c < 3C / 4 \\ 0,5 * A, & \text{otros casos} \end{cases} \quad (3)$$

Función objetivo o calidad de una red: en este caso se utiliza un método envolvente que convierte cada nodo en un clasificador bayesiano y la calidad de la red lo determina la precisión del clasificado correspondiente con una validación cruzada de 10 del conjunto de entrenamiento. De esta forma se identifica la estructura que mejor clasifique representa la mejor solución de todas.

La figura 3 presenta el código general de la propuesta del algoritmo VMO al aprendizaje bayesiano.

```

1 Generación de la malla inicial M de forma aleatoria con p BN
2 Construir una clasificador con cada red y seleccionar la de mejor precisión BNg de M
3 Crear un malla temporal Mtemp ← M
4 Repetir
5   Para cada nodo BNi ∈ M hacer
6     Aplicar ecuación 1 para encontrar sus k vecinos más similares.
7     Determinar el clasificador de mejor precisión entre los vecinos BNi(j)
8     Si BNi(j) tiene mejor precisión que BNi, entonces
9       Aplicar ecuación (BN1UBN2) con BNi(j) y BNi
10      Mtemp ← BN*
11     Fin Si
12   Fin Para
13   Para cada nodo BNi ∈ M hacer
14     Generar un nuevo nodo usando la ecuación 2 con BNg y BNi
15     Mtemp ← BN*
16   Fin Para
17   Seleccionar los dos nodos fronteras de la malla BNj1 y BNj2
18   Generar un nuevo para cada frontera usando la ecuación (BN1\BN2) union (BN2\BN1)
19   Mtemp ← BN*, Mtemp ← BN*
20   Ordenar las BN de Mtemp según su precisión
21   Aplicar operador de limpieza adaptativo a Mtemp
22   Construir M con las p primeras BN de Mtemp
23   En caso que no se cuente con las p BN en M, completar con soluciones aleatorias.
24 Hasta C evaluaciones
    
```

Figura 3. Código general de la propuesta del algoritmo VMO al aprendizaje bayesiano

III RESULTADOS Y DISCUSIÓN

En esta sección se presenta el estudio experimental de la propuesta VMO al aprendizaje estructural (Bayes-VMO). Para ello, utilizamos 14 bases de conocimientos incluidas en el repositorio UCI ML [14], cuya descripción se presenta en la Tabla I donde rd y rc representan cantidad de atributos discretos y continuos respectivamente.

Tabla I. Conjunto de datos y su descripción.

base	rd	rc	clases	casos
mammographic	4	1	2	961
Lung-cancer	56	0	3	32
hepatitis	13	6	2	155
e colic	0	7	8	336
breast-cancer-w	10	0	2	683
contac-lenses	4	0	3	24
hayes-roth-m	4	0	3	132
Monk	1	6	0	2
vote	16	0	2	300
Balance-scale	4	0	3	625
tic-tac-toe	9	0	2	958
Iris	0	4	3	150
Labor	8	8	2	57
Soybean	35	0	19	683

La experimentación comienza realizando un estudio de los parámetros k (cantidad de nodos que determinan la vecindad de cada nodo) y p (cantidad de nodos que integran la malla inicial en cada iteración) del algoritmo. Luego la mejor configuración es utilizada para un estudio comparativo con otros clasificadores bayesianos del estado del arte.

Para el análisis de los resultados se aplica un conjunto de técnicas estadísticas no paramétricas presentadas en [15]. En todos los experimentos se realiza un máximo de 10000 evaluaciones de la función objetivo (valor del parámetro C) y cada variante del algoritmo se ejecutó 50 veces de manera independiente para cada conjunto de datos, utilizándose el valor promedio para el análisis comparativo.

ESTUDIO DE PARÁMETROS

En esta sección, se realiza el ajuste de los parámetros k (cantidad de nodos que determinan la vecindad de cada nodo) y p (cantidad de nodos que integran la malla inicial en cada iteración) utilizando los valores p=(12, 24, 48) y k=(3, 5, 7) respectivamente. La Figura 4 presenta un resumen general de los resultados obtenido por cada variante de estudio en todas las bases de conocimiento, donde se puede apreciar que los mejores resultados se alcanzan con los parámetro p=24 y k=3

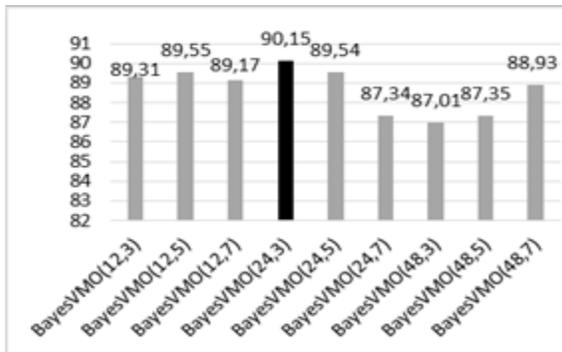


Figura 4. Calidad promedio que alcanzó cada variante en todos los datos

Para determinar si las diferencias son estadísticamente significativas, se aplicó el test de Iman-Davenport [16] con 8 y 104 grados y se obtuvo un valor menor a 0,05 (2,92E-10), determinando que existen diferencias significativas en los resultados computados.

El test de Holm [17], por su parte identificó que los resultados alcanzados para la variante BayesVMO(24,3) mejora significativamente las otras variantes que aparecen en la tabla III. Para el caso de las otras alternativas que no aparecen (BayesVMO(12,3), BayesVMO(12,7), BayesVMO(24,5) y BayesVMO(12,5)), los resultados son similares.

Tabla II. Resultados del test de Holm para BayesVMO(24,3) como control

BayesVMO	z	Valor p	α/i	H
(48,5)	4.209	2.56E-5	0.006 2	R
(48,3)	4.174	2.98E-5	0.007 1	R
(24,7)	3.726	1.94E-4	0.008 3	R
(48,7)	2.863	0.0041	0.01	R

De este análisis se puede determinar que el algoritmo BayesVMO obtiene los resultados más prometedoros con los parámetros p=24 y k=3.

A continuación utilizaremos esta variante para el análisis comparativo con otros clasificadores bayesianos.

ANÁLISIS COMPARATIVO CON OTROS CLASIFICADORES BAYESIANOS

Para el análisis comparativo seleccionamos algunos clasificadores clásicos como: K2, TAN, NB, CBN (clasificador Naive Bayes) y otros más complejos como ByNet, BayesChaid, BayesPSO presentados en [10].

La Figura 5 muestra los resultados promedio de cada clasificador sobre todo el conjunto de datos. Para el caso de los resultados de ByNet, BayesChaid y BayesPSO fueron tomados exactamente como fueron presentados por el autor y se utilizó como condición de parada del algoritmo 10000 evaluaciones de la función objetivo tal y como se presentó en BayesPSO.

Aplicando el análisis estadístico corroboramos que en el grupo de algoritmos existen diferencias en cuanto a resultado (valor de Iman-Davenport 4,316E-4 < 0,05). Por su parte el test de Holm determinó que los resultados alcanzados por nuestra propuesta BayesVMO(24,3) obtuvo valores de clasificación significativamente superiores que los demás clasificadores en estudio. En la Tabla III se pueden observar estos resultados.

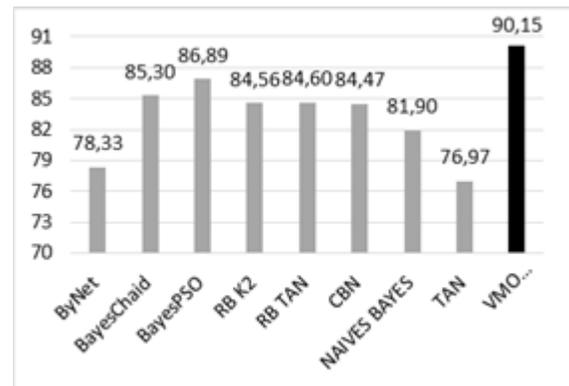


Figura 5. Resultados computados para el análisis com-

Tabla III. Resultado de la prueba de Holms con BayesVMO como algoritmo de control

	z	Valor p	α/i	H
TAN	4.243	2,20E-02	0,0062	R
NAIVES BAYES	3.864	1,11E-01	0,0071	R
ByNet	3.795	1,47E-01	0,0083	R
CBN	2.725	0,0064	0,01	R
RB TAN	2.622	0,0087	0,0125	R
RB K2	2.415	0,0157	0,0166	R
BayesChaid	2.242	0,0243	0,025	R
BayesPSO	2.001	0,0453	0,05	R

IV CONCLUSIONES

Del presenta trabajo se puede concluir lo siguiente:

- Se logró adaptar la meta heurística Optimización basada en Mallas Variables (VMO) para el aprendizaje estructural de un clasificador bayesiano utilizando operaciones de la lógica de conjuntos como la unión y la diferencia.

- En el estudio de parámetro se determinó que el tamaño de la malla y la cantidad de vecino son determinantes para la calidad de los resultados, siendo los valores $p=24$ y $k=3$ los recomendados.

- En el estudio comparativo se demostró que nuestra forma de entrenar estructuralmente redes bayesianas resulto ser competitiva con otras propuestas del estado del arte, alcanzando los resultados más altos de clasificación en los casos de estudio. Resultados que fueron probados estadísticamente.

En un futuro cercano se pretende ampliar nuestra investigación para probar otros mecanismos que permitan mejorar el proceso de exploración el algoritmo VMO y de esta forma encontrar topologías más prometedoras.

V REFERENCIAS

- [1] Chickering, D. (1996). Learning from Data: Artificial Intelligence and Statistics V.
- [2] Friedman, N. G. (1997). Machine Learning. Springer Link, 29-131.
- [3] Spirtes, P. a. (2000). Causation, Prediction, and Search. MIT press.
- [4] Cooper, G. a. (1992). A Bayesian method for the induction of probabilistic networks from data. Machine learning.
- [5] Mosier, C. (1951). The need and means of cross validation. Problems and designs of cross-validation.

Educational and Psychological Measurement.

- [6] Duarte, A. (2007). Metaheurísticas. Librería-Editorial Dykinson.

- [7] Puris, A. a. (2012). Variable mesh optimization for continuous optimization problems. Soft Computing.

- [8] Oviedo, B. a. (2014). Optimización basada en Mallas Variables: Caso de estudio Viajante de Comercio. The 12th Latin American and Caribbean Conference for Engineering and Technology, LACCEI 2014.

- [9] Mark Bartlett, James Cussens, (2017). Integer Linear Programming for the Bayesian network structure learning problem, Artificial Intelligence, Volume 244, Pages 258-271.

- [10] Hui Liu, Shuigeng Zhou, Wai Lam, Jihong Guan, (2017). A new hybrid method for learning bayesian networks. Published in: Journal Knowledge-Based Systems, Volume 121 Issue C, Pages 185-197

- [11] Molina, D. a. (2013). Evolutionary Computation (CEC), 2013 IEEE Congress on: Variable mesh optimization for the 2013 CEC special session Niching methods for multimodal optimization.

- [12] Chávez, M. (2008). Modelos de redes bayesianas en el estudio de secuencias genómicas y otros problemas biomédicos. En M. Chávez, Tesis Doctoral, Universidad Central" Marta Abreu" de las Villas. Villa Clara, Cuba.

- [13] Oviedo, B. a. (2016). Learning Bayesian Network by a Mesh of Points, IEEE World Congress on Computational Intelligence (IEEE WCCI 2016).

- [14] Lichman, M. (2013). Machine Learning Repository. Obtenido de University of California, Irvine, School of Information and Computer Sciences: <http://archive.ics.uci.edu/ml>

- [15] García, S. a. (2009). A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC2005 Special Session on Real Parameter Optimization. Journal of Heuristics, 617-644.

- [16] Iman, D. a. (1980). Approximations of the critical region of the Friedman statistic. Commu. Stat., 571-595.

- [17] Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 65-70.